



UPDATE -

News from the LS User Group

Issue no. 13
February 1996

Contents:	Page
1 Diary	
- LS Workshops	2
- One-day seminar on internal migration	2
- Annual LS User Group meeting on "Health Variations"	2
2 Note on the merger of OPCS and CSO	2
3 Readership survey	3
4 Technical issues:	
- 1991 Small Area Statistics	3
- The 1991 urban-rural indicator	3
- Mapping with Statistical Analysis Software	5
- Travel-to-Work-Areas: latest news	8
- Identifying the same individual at different time points	8
5 Women, occupations and the 1980 reclassification	11
6 STOP PRESS! Seminar on longitudinal findings from the EDP	19

This newsletter is designed to provide rapid dissemination of information on the OPCS Longitudinal Study (LS) and a forum for the exchange of users' views and comments. It is produced by the LS User Support Programme at the Social Statistics Research Unit (SSRU) at City University. All comments and contributions for the newsletter should be sent to Rosemary Creeser, LS Support Programme, SSRU, City University, Northampton Square, London EC1V 0AR tel. 0171 477 8487 Email: rc@uk.ac.city.ssrु. Contributions on IBM -formatted floppy disk are always welcome and should be sent, clearly documented (file name, wordprocessing package and version used) along with a hard copy of the text.



1 Diary

This section highlights forthcoming events of interest to LS Users.

If you are arranging an event and wish to publicise it in future issues of *Update* you should send details to Dina Maher, the LS Administrative Secretary at SSRU.

LS Workshops

SSRU hold regular 2-day workshops. These provide detailed information on the study and enable researchers to gain practical hands-on experience of accessing the data. They are also an ideal opportunity to meet members of the LS Support Team and to discuss the suitability of the LS for exploring specific research questions. The next LS Workshop will be held on 1st/2nd April. A programme and booking form are enclosed.

As part of the hands-on element of the workshop participants are able to specify a statistical analysis of their choice using a small sub-set of variables and a test data-set based on 1% of the LS data. The number of places is limited to ensure that participants get sufficient individual attention and hands-on experience. A fee of £20 is charged to cover documentation and administrative costs. Researchers who plan to start work shortly on projects using LS data, including those planning to use the study for MSc dissertations this academic year, are advised to contact Dina Maher immediately on 0171 477 8486 to reserve a place. (EMAIL: dm@ssru.city.ac.uk)

One-day seminar on internal migration

On Wednesday 3rd April SSRU will be holding a one-day seminar on internal migration. This will take place in the Senate Suite, City University. The seminar will include an update on the migration data available in the LS. There will also be presentations on recent LS work exploring migration between metropolitan and non-metropolitan areas, the changing composition of rural populations and the relationship between migration and social or household change. A non-refundable fee of £45 will be charged to cover administration, a buffet lunch, cheese and wine. A programme and booking form are enclosed. For further details please contact Dina Maher on 0171 477 8486.

LS User Group annual meeting on "Health Variations"

In October SSRU will be holding the annual meeting of the LS User Group at City University, the theme of which is "Health Variations". We hope to include presentations reflecting analyses of event data and the question on limiting long-term illness included in the 1991 Census. The programme for the meeting has not yet been finalised. If you are interested in presenting a paper based on recent LS work, please contact Rosemary Creeser at City University on 0171 477 8487 for further details.

2 Note on the merger of OPCS and CSO

On the 1st April 1996 the Office of Populations, Censuses and Surveys (OPCS) will merge with the Central Statistical Office (CSO) to form a new organisation. This will be known as the Office for National Statistics (ONS). Following the merger, the London offices of ONS will be relocating to a new site in Pimlico, London SW1, towards the end of 1996. Watch this space for details of the new address and telephone numbers.

3 Readership survey

With this issue of *Update* we have included a readership survey. This is part of a review of the newsletters which are currently produced to disseminate information on the LS. How many items did you read in the last issue of *Update*? Are the stories too long/too short/about right? How would you like to receive technical information on the LS? We hope to publish the results of the survey in the next issue of the newsletter. Please use this opportunity to give us your opinion on *Update* and to suggest topics for future issues.

4 Technical issues

4.1 1991 Small Area Statistics, Kevin Lynch

Data from the 1991 Census Small Area Statistics (SAS) are now available on the LS database. These data are a set of statistics using enumeration district as the basic "building block", aggregated to a range of standard areas. The 1991 dataset in the LS contains variables at ward, local authority district and district health authority levels. The 1971 and 1981 SAS data in the LS are only coded to ward level.

The LS includes a total of 145 of the SAS variables covering such topics as age structure, household or ethnic composition and employment. These are linked to LS member records by the appropriate area code of usual residence. Table 1 summarises the 1991 SAS in the LS and shows the areas for which they are available.

The 1971 and 1981 SAS data are coded as proportions of the relevant population (eg proportion of persons in private households aged 0-4). By comparison the 1991 data include the number of individuals or households with particular characteristics and the denominator, and therefore are more flexible.

4.2 1991 Urban-rural Indicator

The variable URRUIND9 has now been added to the 1991 LS members file. This is based on an indicator jointly developed by OPCS, Department of Environment and Ordinance Survey of the urban/rural nature of the LS member's enumeration district of usual residence. It is coded 1=Urban, 2=Rural.

Table 1: Summary of the 1991 Small Area Statistics in the LS

Economic Position:	Male/Female residents aged 16+. Male/Female residents economically active, on government training schemes, unemployed*. (8)
Ethnic composition:	Total residents. Residents in households by ethnic origin. (12)
Wholly moving households:	Number of wholly moving households. (1)
Tenure:	Number of households in permanent buildings. Households by tenure. Owner occupied households. Households with no bath/inside WC. (11)
Cars:	Households with 1/2/3 or more cars. (3)
Rooms:	Number of households. Households with 1,2,7 or more rooms. (4)
Persons/Rooms:	Households with 1-1.5, 1.5 or more persons per room/in non permanent accommodation. (4)

Table 1: Summary of the 1991 Small Area Statistics in the LS (contd)

Marital status/ Economic position	Residents in households over 16. Married male/female residents, economically active married females, students in households. (5)
Age structure:	Residents by age groups. (21)
Sharing: heating. (3)	Number of shared households, shared households with bath/wc/, central heating. (3)
Dependent children:	Number of households with dependent children, 4+ dependent children. (2)
Household spaces:	Total household spaces, with residents, total rooms in household spaces. (2)
Single person: households	Single person households. (1)
Flats/bedsits:	Non self contained flats, bedsits, rooms. (3)
Employment and Industry:	Males and females in employment by industry. (24)
Travel-to-work:	Residents in work by means of travel to work. (10)
Socio Economic Group:	Households by socio-economic group (SEG) of head. (19)

Note: The 1991 SAS relate to the numbers of individuals with particular characteristics in the *area of enumeration*. * Ward, Local authority district or District Health Area.

The figures in brackets () indicate the number of variables.

4.3 Mapping with Statistical Analysis Software, Rosemary Creeser and Kevin Lynch

If you have some experience in using Statistical Analysis Software (SAS) at your own site, either on a mainframe or PC, you may already be familiar with SAS/GRAPH - the SAS module which produces charts, plots, maps and 3-dimensional graphs. We are happy to report that the systems problem that prevented us from using this particular module on the OPCS mainframe computer has recently been resolved. Those of you who have chosen to carry out your LS analyses in SAS may now use the facility to produce a range of high quality colour graphics. The following sequence of questions and answers provide an introduction to one of the most valuable features of SAS/GRAPH - the ability to map data to illustrate the variation in a variable with respect to a given area.

Q: Which of the SAS/GRAPH procedures produces maps ?

The GMAP procedure may be used to produce 2-dimensional (choropleth) and 3-dimensional (block, prism and surface) maps. For most LS work you will find that simple choropleth maps produce the best results. Three dimensional maps can also be used. However, the appearance or "finished result" is much more dependent on the variation in the data.

Q: What type(s) of map are available for use with LS data ?

If you wish to map *individual-level* LS data (eg on the LS member, their family/household or individual household members) currently you may use either region or county level maps. The confidentiality requirements of the LS prohibit the mapping of individual-level data for small geographic areas (eg below county level). However, *aggregate-level* data describing the characteristics of an area, such as the area unemployment rate from the Small Area Statistics, may be used to construct maps showing variation by county district.

Q: What are the minimum requirements to produce a map using GMAP ?

To produce a map using GMAP you need to create a SAS dataset containing an "area identification variable" of the same name, type and length as specified in the *map dataset* you are using.

The SAS dataset containing the response variable or variables you wish to map is referred to as the *response dataset*. The GMAP procedure matches the value of the *response variable* for each unit area in the *response dataset* to the corresponding unit area in the *map dataset* to create the output graphics.

Q: Do I need to specify the response variables in a particular way ?

Not necessarily. If you wish to produce a choropleth (2-dimensional), block or prism map your *response variable* can either be in character or numeric format. For surface maps, however, the response variable must be specified as a numeric variable, with a positive value. GMAP can produce maps for both discrete and continuous variables.

For the SAS example included below, we used the categorical variable PERX as our *response variable*. This was derived in SAS to reflect the percentage of LS members in a county who reported having a limiting long-term illness (LLTI) at the 1991 Census.

Q: Can you give me an example of the SAS code used to produce a simple 2-dimensional map ?

Below we include an example of the SAS code to produce a simple 2-dimensional map illustrating the variation by county in the reporting of limiting long-term illness at the 1991 Census. The map produced using this code is reproduced in figure 1.

Example of SAS code

```
FILENAME XX 'DPLSO.DFOO.CUKL0001'; * OUTPUT FILE FOR MAP; LIBNAME MAPS
'DPLSO.DFOO.CUKLX001.V6DATLIB';

/* Sets the graphic options and specifies how the graphics are to be output. Here we are outputting to a file
in Hewlett Packard Graphics Language (HPGL) format. This may be imported into a Wordperfect
document. */

GOPTIONS DEVICE=HPGL GSFNAME=XX GACCESS=GSASASTD GSFMODE=REPLACE;
/* DAT : Data file produced from LS using Summary to aggregate
to county level. Contains a count of persons with LTILL, total persons,
and percentage with LLTI by county. */

FILENAME DAT 'DPLSO.DP00.CUKL0001.DATLIB(illout)';
```

```
/* Creates a SAS data set and recodes the percentage with a LLTI to quartiles */
```

```
DATA MAPS.COTILL9;INFILE DAT;  
INPUT COUNTY ILL NILL TOT PER;  
IF PER LT 12 THEN PERX=1;  
IF PER GE 12 AND PER LT 13.5 THEN PERX=2;  
IF PER GE 13.5 AND PER LT 15 THEN PERX=3;  
IF PER GE 15 THEN PERX=4;
```

```
/* Defines the legend used for the map */
```

```
LEGEND1  
LABEL=( 'LS Members with LLTI % ' height=10 pct)  
VALUE=(T=1 "<12" T=2"12-13.5" T=3 "13.5-15" T=4 "15+" height=10 pct);
```

```
/* Define patterns for 4 area groupings. As we aimed to reproduce this map  
in monochrome we set all of the colours to white. The value option (V)  
specifies the patterns used eg V=E(mpty), V=M3N045 (heavily shaded pattern of  
parallel lines drawn 450 from the horizontal), V=M3X (a heavily shaded pattern  
of cross-hatched lines) and V=S (a solid pattern). */
```

```
PATTERN1 C=W V=E R=1;  
PATTERN2 C=W V=M3N045 R=1;  
PATTERN3 C=W V=M3X R=1;  
PATTERN4 C=W v=s R=1;
```

```
/* Mapping procedure. Variable COUNTY in both files provides link  
between map and response datasets. The CHORO statement indicates a choropleth  
map is to be produced using the variable PER, which has four levels. */
```

```
PROC GMAP MAP MAPS.LSMAP DATA=MAPS.COTILL9;  
ID COUNTY;
```

```
CHORO PERX/LEVELS=4 LEGEND=LEGEND1;
```

```
/* Use the line below if you not want a legend */
```

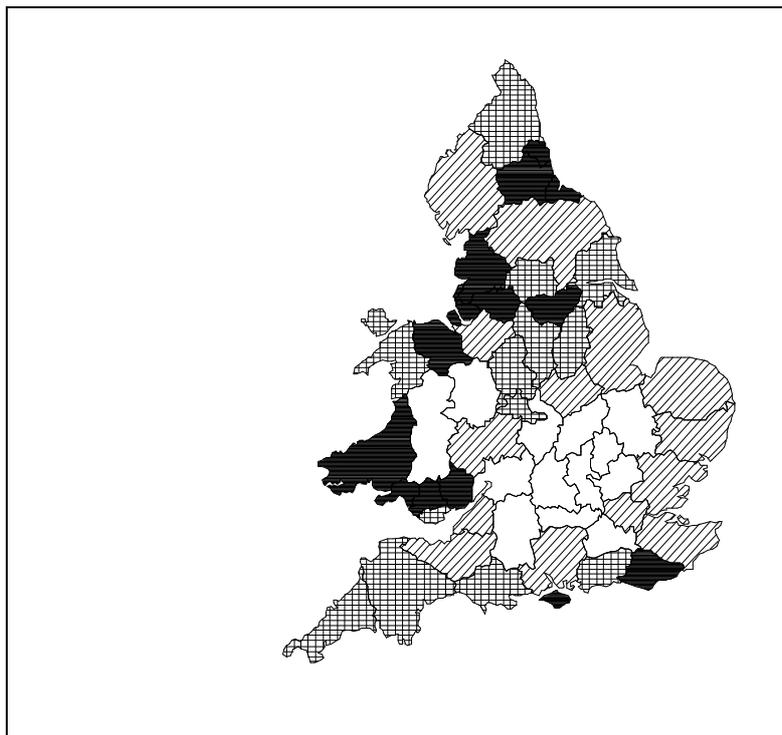
```
* CHORO PERX/LEVELS=4 NOLEGEND;
```

```
/* Use the line below to define the position, font and text for title */
```

```
* TITLE J=CENTER FONT=CENTX 'LLTI By County All Ages';
```

```
RUN;  
QUIT;
```

Figure 1: Map showing the variation in the reporting of limiting long-term illness among LS members enumerated at the 1991 Census, by county



NB: The solid dark areas indicate the counties where at least 15 per cent of LS members reported a limiting long-term illness.

We have recently received a version of SAS GIS. An article on this be included in the next issue of *Update*. In the meantime if you require further information please contact either Rosemary Creeser on 0171 477 8487 or Kevin Lynch on 0171 477 8000 X 4140.

References:

SAS Institute Inc. (1990) SAS/GRAPH Software, Volume 2 - in particular chapter 29 "The GMAP Procedure".

4.4 Travel-to-Work Areas: latest news, Brian Dodgeon

In *Update* no 12, Mike Coombes of Newcastle University gave a full account of the process of defining Travel-to-work-areas (TTWAs), and outlined the current position regarding their possible use with the LS. Three possible approaches were dealt with:

- a. New 1991-based TTWAs produced by Newcastle University and approved by the Central Statistical Office (CSO).
- b. An alternative definition of localities, sponsored by the ESRC and designed by Newcastle University to be of greater interest to the academic community. These areas reflect not only commuting patterns, but also local migration and use of community services such as banks.
- c. The coding of 1991 data in the LS to existing 1981-based TTWAs.

The position on a. is unchanged: we are still waiting for the CSO to sanction the TTWA definitions which Newcastle University produced for approval in Spring 1994.

On b., the news is more optimistic. It is anticipated that the definitions will be available for LS use within a matter of weeks.

As regards c., we are in the process of securing two "look-up" tables from the MIDAS national computing facility at Manchester University, which recodes the 1991 Census data in the following way:

- i. 81 wards to 81 TTWAs
- ii. 91 EDs to 81 wards

We already have the first of these tables and should be receiving the second within the next week or two. We shall then combine them to provide, for each LS member present at the 1991 Census, a code for the 1981 TTWA corresponding to their 1991 enumeration district (ED).

This will enable users to do longitudinal analyses which retain the relative distinctions between two localities which may have been "merged" in 1991 (see Mike Coombes' illustrative example of Sheffield & Rotherham in *Update* no.12, p.20).

4.5 Identifying the same individual at different time points, Judith Wright

In *Update* no.12 some of the technical issues involved in using the LS for intra-household analyses were considered. The article focused on the way relationships between individuals within a household could be found using variables that are the product of census processing and those that have been derived for the LS. Having identified these relationships LS researchers may wish to focus on those members of the household with the same relationship to the LS member at different points in time and find out if they are the same person. These notes illustrate how this type of analysis may be carried out.

Some suggestions for identifying the same individual

LS members are "traced" and "linked" between censuses and vital registrations in a way that members of their census household are not. However, in some cases it is possible to carry out a longitudinal analysis of household members using their date of birth information. Date of birth (day, month, year) is entered on the census schedule for all members of the household. Clearly there will be some misreporting of dates of birth. As a result of reporting or coding errors comparisons based on them at different times may identify a "different person". To allow for the misreporting of dates, researchers may choose to match on two of the three elements of the birth date (day, month, year), and use other variables to substantiate the comparison.

The sex variable may be used to find the right child born to a sample member while marital status may be used in conjunction with dates of birth to identify the same spouse. In the LS, date of birth discrepancies, for example between vital events and census, are "flagged" for LS members but not for the other members of the household in which they are enumerated. This makes it impossible to rule out date of birth discrepancies from a longitudinal analysis of household members.

LS User Guide No 1: Households, families and fertility describes the variables that were derived in the previous computing environment to identify individuals in the same household as the LS member at different time points. Both birth registration data and information from the 1971 and 1981 Censuses were used. These variables include:

1. *Registered child indicators*

These were derived using information on the sex and date of birth of the LS member's first child, as recorded at birth registration and, if the child had not died in infancy, the sex and date of birth of the child enumerated in the LS mother's household at the 1981 Census.

2. Child's father indicator

This was created for children jointly registered at birth. It used the father's date of birth recorded at birth registration and compared it with the date of birth recorded for the male partner living in the household at the 1981 Census.

Equivalent variables, for example identifying a spouse, parent or child of the LS member - may also be derived for the 1991 Census data. A research project which compared the socio-demographic characteristics of LS members by whether they had the same spouse at the 1971 and 1981 Censuses (Ní Bhrolcháin, 1988) is now being extended to spouses of LS members at the 1991 Census.

A set of derived variables are already available in the LS which allow the analysis of the non-members of the household in the LS member's family - such as spouse or parents. However, researchers interested in making longitudinal comparisons need to be particularly careful. Where the "LS member's family" variables are used for identifying the spouse of the LS member, the spouse may not necessarily be the same person at different censuses. To ascertain this further investigation is necessary.

Ongoing Research

A current research project is investigating LS members born in the two intercensal decades 1971-1981 and 1981-1991, to establish their family circumstances at birth registration and their subsequent circumstances at census. (See Clarke and Joshi reference below.) For LS members who are children born in these two decades and living in a family at census the parent(s) recorded at birth registration are being compared with those enumerated in the same family as the LS member.

As I mentioned in *Update* no. 12, in certain circumstances it is difficult to correctly identify the parents of LS members using the LSRELAT variable for 1981 and 1991. For this reason we have only investigated those LS members that were enumerated in "unconcealed families". (An 'unconcealed family' is where the head of the family is also the head of the household.) Finding parents in cohabiting families in 1991 and their equivalent in 1981 - lone parent families with *de facto* spouses present - has involved extra computing to accurately identify parent figures on which to do a comparison of dates of birth.

For each year the parents found at census were compared with those present at registration. We followed the rule mentioned above and assumed that we had the same person if two out of three elements - day, month, year - of the person's date of birth were the same at the two points. New variables were created to count the number of same, different, and missing elements of the dates at each time point. The two out of three rule gave a mutually exclusive set of possible outcomes - "different parent", "same parent" and "insufficient information on which to test whether the parent was the same or not". In order to see the scale of the difference we further split these into "definitely different" (three out of three different) and "probably different" (two out of three different).

The outcomes for the date of birth comparison for each parent were concatenated to produce outcomes for both parents. These ranged from "both parents same", through "mother same, father definitely different", "mother same, father possibly different" to "mother same, insufficient information on father" etc. There were sixteen possible combinations of whether one or both parents were the same at census as at birth registration.

A further refining exercise was carried out in order to clear up the cases where it was difficult to conclude whether the parents were the same or not. For this, the sixteen "same parents" outcomes were cross tabulated with the eight different combinations of registration type to census situation and the marital status of each parent at census (see tables 2 and 3).

Each of these "Registration to Census" possibilities were crosstabulated against mother's and father's marital status at census. For example, if a birth had been registered within marriage but the LS member was enumerated at the 1991 Census in a cohabiting couple family, and the mother had the same dates of birth at census and at birth registration but experienced a change in marital status, it is likely that a "possibly different father" was definitely a different father to that at birth registration.

These tables allowed rules to be drawn up for allocating particular combinations of marital status of parents and their census and registration situation to make decisions about the likelihood of the parent(s) being the same or not. Although possibly only re-allocating a small proportion of ambiguous cases the process took advantage of the added value in the LS of combining variables about relationships between members of the household, about the individuals in the household and changes over time.

Table 2: First decade births to 1981 comparison:

At birth registration	by	At Census
Married - first)	Married couple family
Married - remarried (mother only))	Lone parent with <i>de facto</i> spouse
Outside marriage - sole)	Lone parent family
Outside marriage - joint)	

Table 3: Second decade births to 1991 comparison:

At birth registration	by	At Census
Married - first)	Married couple family
Married - remarried (mother only))	Co-habiting couple
Outside marriage - sole)	Lone parent family
Outside marriage - joint)	

Results showed that of those children born in the first intercensal decade (1971-1981) and enumerated in a married couple family at the 1981 Census, 93.3 per cent were living with the same parents at census as at birth registration. Of those children born between 1981 and 1991 and living in a married couple family in 1991, the percentage with the same parents at the two time points was 94.5 per cent.

The exercise of finding parents who were *de facto* spouses of lone parents in 1981 meant that a further 1505 cases were found where both parents could be compared. Of these 45.4 % (683) did have the same parents at registration as at census. This indicates the value of being able to identify a cohabiting couple as a unit of analysis in the LS in 1991.

Other relationships and individuals may be identified in a similar way. The main issues to consider are how to correctly identify those relationships and the criteria by which to decide whether a match has been made. Results from projects using this type of analysis will be reported in future issues of *Update*.

References:

Clarke L and Joshi H (forthcoming 1996) *Living arrangements of children*

Ní Bhrolcháin M (1988) "Changing partners: a longitudinal study of remarriage", *Population Trends*, 53, pp 27-34

Penhale B (1990) *LS User Guide no. 1: Households, families and fertility*, London: LS Support Programme, SSRU, City University

5 LS Research

Women, occupations and the 1980 reclassification, Louisa Blackwell, SSRU, City University

Summary

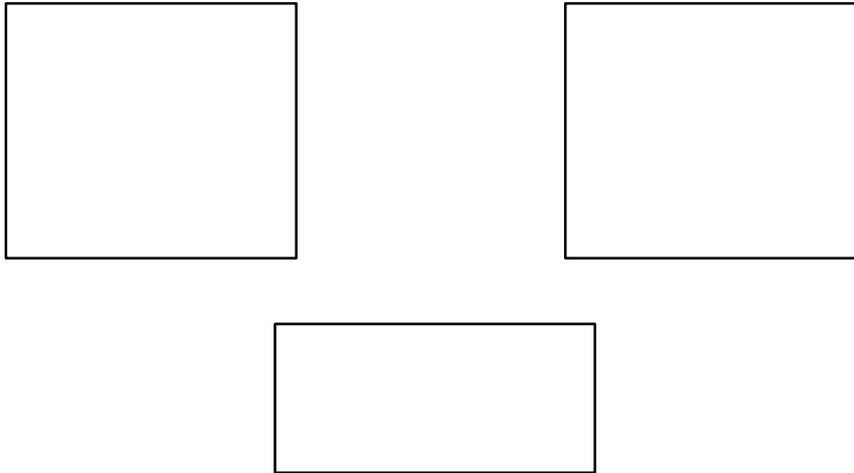
Occupational sex segregation, or the tendency for women and men to work in different occupations, is a long-established and world-wide phenomenon. It is often held responsible for the persistence of the wages gap, which has survived the implementation of equal pay legislation. Because occupational segregation appears to be universal and particularly enduring in its nature, any differences in the level of segregation either over time or place are of particular theoretical interest. However, analyses of time series and longitudinal data are often complicated by changes in the framework used to classify workers by occupation. Typically, researchers have reported their research findings with a cautionary note about the possible impact that occupational reclassification may have had, though without quantification (for example see Dex, Joshi and Macran, 1996). This article investigates the effect of the 1980 occupational reclassification carried out by OPCS. The analyses reported here were carried out as part of a Ph.D dissertation on occupational sex segregation. One of the objectives of these analyses is to be able to adjust the occupational data in the LS for artefactual differences arising from changes in the classification system.

5.1 Background

When comparing the linked occupations of LS members, differences in the 1971 and 1981 sex ratios of members' occupational groups may arise for a number of reasons. Firstly, individuals may have actually changed occupation. For example women may have switched from working in "female"-typed occupations, that is occupations in which women were over-represented at both censuses, to occupations which were disproportionately "male" at both censuses (Type 1 differences). Often these "real" occupational shifts are of most interest to researchers. A second type of difference would arise if an individual, or group of individuals, continued in the same job but because of structural change in the inter-censal period, the sex ratio in their occupation changed so that they were nudged into a different occupational concentration category (Type 2 differences). Those "integrated" occupations which are close to the boundaries between "male" and "female" occupations are more susceptible to this sort of movement. The third possibility, which is the main concern of this article, is that observed changes in patterns of occupational concentration have arisen as a result of occupational reclassification, or the re-grouping of jobs (Type 3 differences), and are therefore artefactual. This typology of differences does not cover all of the inconsistencies that may interfere with longitudinal data comparisons, but differentiates and disentangles these three effects. Type 3 differences need to be quantified and filtered out of the data so that differences in segregation patterns, that is type 1 and type 2 differences, in their "uncontaminated" form, can be analysed.

The 1980 reclassification

The 1971 Census used the 1970 Classification of Occupations (OPCS, 1970) which involved 223 occupational titles. By comparison the 1981 Census adopted an occupational classification system which in its basic form used 351 titles, which could be further expanded to 549 job titles, described in the 1980 Classification of Occupations (OPCS, 1980). To show the comparability between the two, OPCS double-coded a one per cent sample of the 1971 Census returns for persons in employment, both full-time and part-time, using both classification systems. This produced two cross-tabulations of 1971 by 1981 occupational codes - one for women and one for men. By analysing these cross-tabulations it has been possible to isolate the effect that reclassification has on the LS results. The process is summarised in figure 2. (Please note that this is not available in the .PDF version of *Update*.)



The OPCS double-coded sample and the Warwick Conversion Programme

Peter Elias and colleagues at the Institute for Employment Research, Warwick University, have facilitated the analysis of these cross-tabulations by putting them into a machine-readable form in their "Warwick Occupational Conversion Programme". This is available via the ESRC Data Archive. The programme produces groupings of occupations related to the 1980 Classification of Occupations which are consistent with user-defined groupings based on the 1970 classification. The 1970/80 conversion is achieved on the basis of the best statistical match between Occupational Unit Groups (OUGs) in the two classifications. For the analyses described in this article the Warwick programme was adapted to cross-tabulate two sets of user-defined groupings, based on the percentages of female workers in the OUGs in the 1970 and 1980 classifications in the double-coded sample.

Method

For each classification, workers were re-grouped into ten occupational concentration categories. This regrouping was achieved by allocating to each occupation a "percentage female" based on the proportion of women in that group in the double-coded sample. Cross-checking these percentages with those in the 10 per cent published tables for the 1971 Census revealed that, for over 95 per cent of the double coded sample, there was less than a one per cent difference in the percentage female for 202 occupational groups. A difference of more than 5 per cent was found for just two occupational groups, representing just over 1 per cent of the double-coded sample. On the basis of this the double-coded sample may be considered reliable in terms of its representativeness of the sex composition of OUGs. These percentages were then categorised into ten per cent bands, or "ratio groups".

The final output from the programme was a 10 by 10 matrix. The rows of this correspond to the 1970 ratio groups, while the columns represent the 1980 ratio groups. Cells on the main diagonal represent those workers who stayed in the same ratio group through reclassification. Entries off the main diagonal show movement arising from reclassification. It is important to highlight that none of these workers actually changed jobs, but have had the occupational map redrawn around them by the reclassification.

Errors in the double-coded sample

In the coding of the one per cent sample, occupations could have been allocated to the wrong Occupational Unit Groups due to coding errors. Coders may introduce both random and correlated errors to occupational data. The former are caused by mistakes in the allocation of codes to occupational groups, while correlated errors arise from systematic differences in the way that individual coders apply the occupational classification. This type of error is more problematic in office-based coding systems, which OPCS used until the mid-1980s, because they rely on fewer individual coders. Jean Martin et al (1995) found that when office-based coding was compared with a sample which had been "expertly" coded, the level of agreement was 80 per cent. Collapsing coding frames to produce socio-economic groups and social classes raised the level of agreement to 90 per cent and 89 per cent respectively.

In the analysis reported here the occupational classification is collapsed down to produce ten ratio groups, reflecting the percentage of female workers. These groups are not merely an aggregation of codes but take on meaning in a similar way to the socio-economic groups and social classes mentioned above. Therefore the extent of coding error should be reduced. However, entries off the main diagonal indicate the extent of miss-matching arising from the reclassification of occupations. These will be slightly compounded by coder errors. An attempt has been made to adjust for these errors below.

5.2 The effect of reclassification on Occupational Unit Groups

It is not always easy to track Occupational Unit Groups across the reclassification. Both the 549 and the 351-title versions of the 1980 Classification of Occupations were more detailed than the 1970 predecessor. This article focuses on the 351-title 1980 classification as this is the version used in the LS. (An analysis of the effect of reclassification to the 549 category scheme has also been carried out, the results of which will be reported in my final dissertation.) 1970/1980 differences in the definitions of unit groups mean that single 1970 codes almost always become dispersed across several 1980 codes. Similarly 1980 codes can be seen to be composed of several 1970 groups.

Table 4 lists the five largest occupational groups in the 1970 classification, showing their gender composition and ranking with "male" and "female" occupational categories. Table 5 shows how these occupational groups became dispersed over the 1980 classification scheme. This focuses on those destinations to which at least one per cent of the 1970 group was allocated.

In each case, 1970 OUGs were dispersed over several 1980 groups. This illustrates the absence of 1:1 matching and the extent of inconsistency between the two classifications. The matrices, produced by the Warwick Programme, reproduced in tables 6 and 7, illustrate the full extent of ratio group dispersion generated by the 1980 reclassification.

Table 4: The five largest occupational groups in 1971, coded to the 1970 classification.

Code & title	Total no. workers	% fem*	No. women (rank)	No. men (rank)
139 Clerks, cashiers	222345	62	138710 (1)	83635 (1)
144 Shop salesmen & assistants	83794	80	67183 (3)	16611 (23)
141 Typists, s/hand writers & secretaries	71354	99	70479 (2)	875 (193)
143 Proprietors & managers, sales	63552	33	20728 (9)	42824 (3)
122 Drivers of road goods vehicles	51833	2	1241 (70)	50592 (2)

* Percentages drawn from the double-coded 1971 sample

Shifts in the pattern of occupational concentration: the 10 by 10 matrices

Table 6 cross-tabulates the ratio groups for men arising from the 1970 classification with those derived from the 1980 classification. The percentages on the main diagonal, which are highlighted, represent a "no change" position. Reclassification did not alter the ratio group for workers in these occupations.

The ratio groups of 77 per cent of men were unaffected by the reclassification. For the remaining 23 per cent, the percentage female was sufficiently altered to force their occupations into a different ratio group. Most of the shifts off the main diagonal were to one of the adjacent groups. For example, 0.5 per cent of all men shifted from ratio group 5 (GT 40 and LE 50 per cent female) in the 1970 classification to ratio group 6 (GT 50 and LE 60 per cent female) when the 1980 classification was used. The most significant shifts occurred at the male end of the occupational spectrum which is not surprising given that men are very highly concentrated in these types of occupations.

Table 7 shows the 10 by 10 matrix for women. Twenty three per cent of women were allocated to a different ratio group as a result of reclassification. Most of the movement was from the 1970 ratio groups where women are most heavily concentrated. Because there are fewer OUGs in very feminised occupations, the off-diagonal outliers are more easily accounted for. For example, most of the women in cell 2:8 (1970:1980) had changed ratio groups because the 1980 classification aggregated cleaners of offices, homes and streets with those working in manufacturing. In the 1970 classification cleaners in manufacturing were grouped with "other labourers not elsewhere classified".

Given the possibility of coder errors, the upper limit to type 3 error in these tables is 23 per cent. This could lead to substantial overstatement of longitudinal change if the variable is measured at the 10-fold level but reclassification does not often generate big shifts.

Table 5: Distribution of the five largest 1970 occupational groups across the 1980 classification (351-title version)

1970 code	1980 code	1980 Occupational title	per cent fem**	% men*	% women*
139 (62% fem)	115	Clerks & cashiers	61	92.9	88.2
	116	Retail shop cashiers	97	***	3.9
	117	Receptionists	98	***	4.2
	Total no. Workers			83635	138710
144 (80% fem)	125	Shop salesmen and assistants	82	85.4	95.9
	127	Petrol pump, forecourt assistants	52	6.1	1.7
	133	Sales representatives	3	1.0	***
	333	Storekeepers	12	1.4	
	Total no. Workers			16611	67183
141 (99% fem)	24	Officials of trade associations	14	5.8	
	29	Managers' personal assistants	41	26.1	***
	115	Clerks and cashiers	61		1.7
	118	Typists, shorthand writers and assistants	99	58.0	96.5
	Total no. Workers			875	70479
143 (33% fem)	15	Buyers (retail)	45	1.6	2.8
	16	Buyers (not retail)	8	1.1	***
	101	Proprietors & managers (sales)	31	86.9	90.2
	128	Roundsmen, van salesmen	6	1.8	***
	131	Scrap dealers	7	2.9	***
	Total no. Workers			42824	20728
122 (2% fem)	326	Drivers of road goods vehicles	2	96.5	95.0
	327	Other motor drivers	5	1.4	5.0
	Total no. Workers			50592	1241

* percentages apply to the original occupational group in the 1970 classification

**These percentages were calculated using the distributions in the 1 per cent double-coded sample.

*** less than 1 per cent, possibly none, workers in this group.

Table 6: The 1970 and the 1980 cross-classification: 10 by 10 ratio groups matrix for men

		1980										
		1	2	3	4	5	6	7	8	9	10	
1970	1	51.86	0.97	0.89	0.41	0.11	0.01	0.09	0.11	0.15	0.00	54.61
	2	7.04	6.09	0.32	0.17	0.17	0.00	0.24	0.36	0.00	0.01	14.40
	3	0.61	2.65	2.91	0.45	0.14	0.09	0.08	0.02	0.01	0.00	6.97
	4	0.58	0.58	1.41	4.57	0.50	0.14	0.19	0.08	0.01	0.00	8.07
	5	0.47	0.10	0.12	0.14	0.66	0.50	0.09	0.31	0.08	0.00	2.47
	6	0.10	0.00	0.01	0.16	0.18	0.21	0.02	0.00	0.00	0.00	0.68
	7	0.19	0.06	0.05	0.01	0.38	0.02	7.53	0.04	0.00	0.02	8.31
	8	0.07	0.02	0.24	0.02	0.07	0.00	0.07	1.22	0.01	0.00	1.72
	9	0.02	0.03	0.01	0.01	0.00	0.09	0.01	0.46	1.60	0.00	2.22
	10	0.01	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.02	0.48	0.54
		60.97	10.51	5.97	5.92	2.22	1.07	8.34	2.61	1.88	0.51	100.00
TOTAL ON THE DIAGONAL=77.1												
TOTAL NO. MEN = 1,322,115												

Table 7: The 1970 and the 1980 cross-classification: 10 by 10 ratio groups matrix for women

		1980										
		1	2	3	4	5	6	7	8	9	10	TOTAL
1970	1	2.09	0.21	0.15	0.18	0.11	0.04	0.07	0.12	0.13	0.00	3.10
	2	0.86	1.50	0.18	0.11	0.20	0.00	0.35	0.58	0.02	0.25	4.05
	3	0.16	0.99	1.60	0.50	0.19	0.15	0.19	0.04	0.13	0.10	4.05
	4	0.17	0.22	1.00	4.04	0.60	0.31	0.71	0.27	0.09	0.04	7.45
	5	0.04	0.04	0.08	0.14	0.93	0.95	0.38	0.61	0.32	0.25	3.76
	6	0.11	0.01	0.00	0.32	0.32	0.52	0.08	0.06	0.00	0.14	1.55
	7	0.01	0.04	0.04	0.01	0.71	0.10	21.58	0.13	0.13	1.42	24.17
	8	0.00	0.01	0.18	0.00	0.13	0.00	0.23	6.73	0.05	0.04	7.36
	9	0.01	0.07	0.05	0.01	0.01	0.15	0.05	4.77	13.91	0.16	19.19
	10	0.01	0.00	0.00	0.01	0.11	0.00	0.23	0.20	0.29	24.49	25.33
TOT	3.47	3.09	3.28	5.31	3.31	2.21	23.86	13.50	15.08	26.89	100.00	
TOTAL ON THE DIAGONAL= 77.4												
TOTAL NO. WOMEN =758,750												

Key:

- 1 (GT 0 and LE 10 per cent female), 2 (GT 10 and LE 20 per cent female),
 3 (GT 20 and LE 30 per cent female), 4 (GT 30 and LE 40 per cent female),
 5 (GT 40 and LE 50 per cent female), 6 (GT 50 and LE 60 per cent female),
 7 (GT 60 and LE 70 per cent female), 8 (GT 70 and LE 80 per cent female),
 9 (GT 80 and LE 90 per cent female), 10 (GT 90 and LE 100 per cent female)

Reclassification and the distributions of women and men across "male", "mixed" and "female" occupations

Comparing the distributions of women and men in "female", "mixed" and "male" occupations is becoming increasingly popular as a way of measuring change in gender segregation and reflects contemporary theoretical concern with "integrated" occupations rather than the traditional dichotomy of "male" and "female" occupations (Hakim, 1993).

This three-fold categorisation of occupations is relatively simple to interpret. Tables 8 and 9 were devised to facilitate comparisons of the gendered occupation shifts in the LS, using the 10 by 10 matrices in tables 6 and 7. For the purpose of this analysis, "female", "mixed" and "male" occupations have been defined using boundaries at 70 and 30 per cent female. (An occupation is deemed to be 'female' or 'male' if either sex account for more than 70 per cent of the work force. 'Mixed' occupations are those where the percentage of female and male workers is closer - eg 45 per cent female and 55 per cent male.)

Table 8: The effect of reclassification on men in "male", "mixed" and "female" occupations

MEN	MALE	MIXED	FEMALE	TOT(%)
MALE	73.36	1.96	0.67	75.98
MIXED	3.69	15.3	0.55	19.53
FEMALE	0.40	0.30	3.79	4.48
TOT(%)	77.44	17.55	5.00	100.00

TOTAL ON THE DIAGONAL= 92.4%

Table 9: The effect of reclassification on women in "male", "mixed" and "female" occupations

WOMEN	MALE	MIXED	FEMALE	TOT(%)
MALE	7.74	2.08	1.37	11.20
MIXED	1.78	31.69	3.46	36.92
FEMALE	0.33	0.92	50.64	51.89
TOT(%)	9.84	34.69	55.47	100.00

TOTAL ON THE DIAGONAL= 90.1%

By using fewer, broader occupational categories, occupational miss-matching as a result of reclassification has been reduced. Aggregating ratio groups in this way has increased the main diagonal totals to 92.4 per cent for men and 90.1 per cent for women. These tables can be used to adjust the 1971/1981 LS results. Cell adjustments can now be made to filter out artefactual shifts in occupational ratio groups arising from reclassification.

The double-coded sample includes some degree of coder error - the effect of which is to generate some artefactual miss-matching in the table. This arises not because the 1970-code and 1980-code ratio groups do not agree, but because workers have been coded to the wrong occupational group in either classification. Consequently, these tables needed further adjustment. Aggregating the 371 SOC codes with employment status information to produce social class codes reduced the level of coder error from 20 per cent to 11 per cent. The aggregation of occupations into "male", "mixed" and "female" groups reduces the

effect of coder error even further, especially given the concentration of OUGs within "male" occupations (see table 10).

Table 10: Distribution of OUGs across "male", "mixed" and "female" categories

Classification		MALE	MIXED	FEMALE	Total
1970	No	155	43	24	222
	%	70	19	11	100
1980 (351-title)	No	244	63	41	348
	%	70	18	12	100

If the amount of miss-matching in the 3 by 3 matrices arising from coder error is assumed to be 3 per cent, tables 8 and 9 can be adjusted to produce tables 11 and 12. Adjustments have been made to preserve cell proportions within the group of cells on the main diagonal and within the group of cells off the main diagonal. At the three-fold level reclassification appears to introduce artefactual change in between 5 and 7 per cent of cases. These tables can now be used to make net adjustments to tables produced using longitudinal data.

Table 11: The effect of reclassification on men in "male", "mixed" and "female" occupations, adjusted for coder error

MEN	MALE	MIXED	FEMALE	TOT(%)
MALE	75.62	1.22	0.42	77.26
MIXED	2.29	15.77	0.34	18.41
FEMALE	0.25	0.19	3.90	4.34
TOT(%)	78.17	17.17	4.66	100.00

TOTAL ON THE DIAGONAL= 95.3%

Table 12: The effect of reclassification on women in "male", "mixed" and "female" occupations, adjusted for coder error

WOMEN	MALE	MIXED	FEMALE	TOT(%)
MALE	7.98	1.50	0.99	10.47
MIXED	1.28	32.67	2.49	36.43
FEMALE	0.24	0.66	52.20	53.10
TOT(%)	9.49	34.83	55.68	100.00

TOTAL ON THE DIAGONAL= 92.9%

Conclusions

The matrices produced by this analysis enable LS researchers to adjust 1971/81 results to take account of the effect of the 1980 reclassification. Although the 1980 reclassification was quite radical, with little 1:1

matching of OUGs in the two classifications, it produced very little change in gender concentration in the occupational structure.

This analysis, which was conducted as part of a Ph.D into occupational sex segregation, has also generated a reliable distribution of the 1971 labour force classified to the 1980 classification. This may be used to facilitate analyses of the changes which took place between 1971 and 1981 in the sex composition of OUGs.

Finally, it is important to remember that the one per cent sample relates to the population of England and Wales as a whole. The observed patterns would not necessarily be the same for particular age-groups or geographical areas.

References:

Dex S, Joshi H and Macran S (1996) "A Widening Gap between Britain's Mothers", *Oxford Review of Economic Policy* (in press)

Hakim C (1993) "Segregated and Integrated Occupations: A new approach to analysing social change", *European Sociological Review*, 9.3, pp 289-314

Martin J, Bushnell D, Campanelli P and Thomas R (1995) "A comparison of interviewer and office coding of occupations", *OPCS Survey Methodology Bulletin*, 37: London: HMSO pp, 21-27

OPCS (1970) *Classification of Occupations 1970*, London: HMSO

OPCS (1980) *Classification of Occupations 1980*, London: HMSO

Thomas R (1986) "Classification of Women's Occupations", *OPCS Survey Methodology Bulletin*, 18, London: HMSO, pp 5-16

Thomas R and Elias P (1989) "Development of the Standard Occupational Classification", *Population Trends*, 55, London:HMSO, pp 16-21

6 STOP PRESS!

Seminar on longitudinal findings from the EDP, a French equivalent to the LS

On May 13th at 3.30pm Dominique Rouault from INSÉE (Paris) will be giving a seminar on recent research findings from the Echantillon Démographique Permanent (EDP) at the Social Statistics Research Unit, City University. This will cover the following areas of work:

"Youth and the transition to work: evidence of inter-generational patterns" **and**

"Electoral participation: social exclusion and voluntary abstention" .

If you wish to attend please contact Dina Maher on 0171 477 8486 to arrange a day-pass.

We will be delighted to receive your views and comments on any of the articles in this newsletter.